

Article

Not peer-reviewed version

Mapping Thematic Clusters with the IEEE Controlled Vocabulary: A Bibliometric Analysis of Energy Systems Technology

[Boris Chigarev](#) *

Posted Date: 5 November 2025

doi: 10.20944/preprints202511.0308.v1

Keywords: controlled vocabulary; IEEE Terms; terms in titles and abstracts; term co-occurrence map



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Mapping Thematic Clusters with the IEEE Controlled Vocabulary: A Bibliometric Analysis of Energy Systems Technology

Boris Chigarev

Oil and Gas Research Institute, Russian Academy of Sciences, Moscow, Russia; bchigarev@ipng.ru

Abstract

Knowledge transfer between different fields, particularly through thematic analysis, is essential for identifying new research topics. Identifying gaps in research is crucial, as it highlights the importance and relevance of initiating new scientific research. It is advisable to use a controlled subject vocabulary to describe research topics, especially when the frequency of terms needs to be extracted from titles and abstracts. The aim of the study was to analyze the use of the expanded IEEE controlled vocabulary in forming thematic clusters by VOSviewer, using bibliometric data from the IEEE Xplore database related to energy systems technologies. It included a comparison of results obtained based on different sets of IEEE terms and author keywords. This analysis aimed to identify potential issues associated with using a controlled vocabulary to identify research gaps, as well as to explore the potential of this methodology beyond the IEEE subject area. An important aspect of the study was understanding the optimal use of terms derived from the controlled vocabulary. The study suggests that the most comprehensive dictionary of terms may not be the most effective tool for identifying relevant topics. This indicates that the use of exclusively technical methods for selecting terms in research dictionaries should be avoided. Although creating a dictionary can greatly assist experts in analyzing their areas of expertise, it is essential that the dictionary undergoes thorough expert evaluation to ensure its relevance and effectiveness for a specific analytical task. Furthermore, the concept of a self-sufficient analytical method is somewhat of a fallacy; while analytics can assist in making informed decisions, it does not replace the need for expert knowledge in the analytical process. This work is preliminary in nature and involves conducting a follow-up analysis of opportunities for transferring knowledge from the IEEE field to the oil and gas sector. The goal will be to identify existing gaps in oil and gas research that can be filled with knowledge from the IEEE field.

Keywords: controlled vocabulary; IEEE terms; terms in titles and abstracts; term co-occurrence map

Introduction

Brief Literature Review

This section discusses the significance of the research topic presented in this paper based on a review of the relevant literature. It examines various characteristics of previous studies and highlights the rationale for choosing the specific objective of this study, emphasizing its importance in the context of existing scientific works.

The following scholarly publications discuss methods for identifying themes and transferring knowledge between fields, often using thematic analysis and lexical approaches. They discuss the importance of identifying gaps in research in order to discover new and relevant research topics in the recipient field of knowledge.

The article [10.32996/ijllt.2022.5.12.6] discusses software developed for thematic analysis based on a manually created dictionary containing approximately 5,500 thematic words. It highlights the

effectiveness of using dictionaries in text analysis, allowing words to relate to multiple themes. This methodology helps identify key themes by determining the most frequently occurring thematic words detected by the software.

It should be noted that the approaches currently used to analyze scientific and technical texts and search for connections between different fields of knowledge have long been used by linguists, albeit not at such a computerized level. An illustration of this can be found in the work [10.1016/0304-422X(94)90011-6]. In this work, language is viewed as a window into the world of the mind. Language is also a window into the world of culture. By analyzing texts, one can study the interaction between human cognition and culture. By analyzing texts, one can describe the cognitive similarities and differences between people that form the basis of culture. Such analysis allows us to identify similarities and differences between cultures, as well as changes within cultures. This article explores the relative advantages of using content analysis and cartographic analysis to extract and analyze cultural characteristics based on a set of texts.

As an example of contemporary research, consider the article [10.1371/journal.pone.0329302], which analyzes the importance of interdisciplinary exchange in the development of science and the problems of tracking knowledge flows in interdisciplinary fields. It presents a new network analysis framework that uses citation data to study the dynamics of knowledge transfer. By applying dynamic community detection to evolving citation networks, this framework identifies research areas and maps interdisciplinary integration, revealing gaps in knowledge. This work aims to support strategies for synthesizing ideas.

The bibliometric and scientometric approaches to the issue under consideration have a long history, as demonstrated by the following publications. Interdisciplinarity can manifest itself in the use of concepts from different disciplines by individual scientists. The paper [10.1007/s11192-009-0121-z] focuses on how the convergence of ideas contributes to new discoveries. It uses science maps to identify potential interdisciplinary connections, analyzed using contextual analysis of co-citations.

Scientists acquire and transfer knowledge daily, though many flows of this knowledge are often unobserved. The review [10.1007/s11024-024-09542-2] examines the application of bibliometric methods in studying these flows, highlighting that traditional bibliometrics has primarily focused on formal knowledge through citation data and is now shifting towards informal knowledge within social networks. However, studies on interpersonal knowledge flows remain limited. The review emphasizes the untapped potential of bibliometric methods and proposes directions for future methodological advancements.

Comment: “network analysis framework that uses citation data” can be modified to “network analysis using controlled vocabulary terms.”

Knowledge transfer can also occur through the implementation of achievements in rapidly developing areas of research into a specific subject area. Here are some examples of such works.

The [10.2118/226792-MS] article discusses the transformative impact of advanced digital technologies, including big data analytics, artificial intelligence, and automation, on petroleum engineering. It emphasizes the importance of integrating traditional domain knowledge with digitalization, identifying this combination as crucial for success. The interaction between deep domain expertise and data-driven approaches enhances established engineering practices, leading to improved operational efficiency and promoting sustainable growth. Key technologies highlighted include AI-based modeling, machine learning for predictive maintenance, and real-time data visualization platforms.

The oil and petrochemical industries face pressures to enhance efficiency, minimize environmental impacts, and respond to evolving market dynamics. The [10.36713/epra24128] study employs a thorough literature review to explore the development of Process System Engineering (PSE) from traditional optimization methods to the incorporation of artificial intelligence (AI) through hybrid modeling approaches. It considers empirical data from diverse regions, including the United States, Iran, and China, bolstered by quantitative market analytics and performance metrics. Findings indicate that the integration of machine learning (ML) with PSE can effectively address the

limitations present in purely data-driven or non-physics-based methodologies by leveraging physics-informed technologies.

Pressure–volume–temperature (PVT) properties of crude oil are considered the most important properties in petroleum engineering applications as they are virtually used in every reservoir and production engineering calculation. Determination of these properties in the laboratory is the most accurate way to obtain a representative value, at the same time, it is very expensive. However, in the absence of such facilities, other approaches such as analytical solutions and empirical correlations are used to estimate the PVT properties. The study [10.1115/1.4050579] demonstrates the combined use of two machine learning (ML) technique, viz., functional network (FN) coupled with particle swarm optimization (PSO) in predicting the black oil PVT properties such as bubble point pressure (Pb), oil formation volume factor at Pb, and oil viscosity at Pb. This study also proposes new mathematical models derived from the coupled FN-PSO model to estimate these properties.

Many oil and gas companies are exploring opportunities in the renewable energy market, particularly in subsea large-scale hydrogen storage systems, which play a crucial role in the global energy transition. Companies leverage their subsea oil and gas expertise to create these hydrogen storage solutions and emphasize the importance of efficient knowledge transfer to integrate new stakeholders into the development process [10.1002/iis2.13076].

Production optimization is vital for closed-loop reservoir management, focusing on maximizing economic benefits through optimal development schemes. Traditional methods often operate in isolation, neglecting knowledge from past optimizations, leading to a high number of simulations that can be computationally expensive. The paper [10.2118/219732-PA] introduces a competitive knowledge transfer method designed to utilize insights from previously solved tasks to improve production optimization outcomes.

The transfer of knowledge between broader and specialized fields is essential for sustainable technological development. Bibliometric studies addressing this knowledge transfer can offer comprehensive insights into the issue.

The article [10.1007/s13132-025-02814-6] discusses that knowledge transfer has been identified as one of the key factors for innovation and business success in the knowledge economy, and how networks are strategic tools that can contribute to this. Performance analysis, scientific mapping, and dynamic network analysis were performed using VOSviewer and SciMat software. This study identifies the most influential participants in this field, presents the evolution of the research direction, describes thematic clusters, and outlines a research program.

The study [10.1007/s10961-019-09774-5] examines the field of technology transfer (TT) in academic research, noting its significant growth and increasing scientific interest in recent years. Using a bibliometric approach, the study examines current issues and their interrelationships in order to identify influential topics and suggest directions for future research. It analyzes the co-authorship network to assess the existing literature, determines the current state of research in the field of TT, and identifies five main areas of research along with related topics.

Organizations rely on external sources to acquire knowledge, especially in the context of university-industry (U-I) knowledge transfer. This topic is becoming increasingly important as U-I collaboration fosters the innovation that companies need to remain competitive. The purpose of the article [10.1108/VJKMS-07-2024-0270] was to examine this topic in detail by conducting a bibliometric systematic review of the literature with a focus on knowledge transfer, collaboration, interaction, and interdependence of new research areas.

Despite the importance of knowledge transfer for successful cross-border acquisitions, contemporary literature does not pay sufficient attention to knowledge management in the acquired company. The study [10.1108/JKM-04-2024-0494] is the first to examine the dynamics of knowledge transfer after an acquisition, focusing on knowledge integration, learning capacity, and knowledge reverse transfer. The authors aim to analyze existing research on these topics and suggest directions for future research in the field of cross-border acquisitions.

Research in the field of knowledge transfer (KT) has attracted considerable attention from scholars over the past decade, although bibliometric and visualization studies remain limited. The study [10.1109/ACCESS.2021.3061576] analyzes key trends in the KT literature using data from the Scopus database, offering a comprehensive overview of KT trends and trajectories through visual diagrams. This approach aims to help researchers and practitioners understand current trends and identify future research directions.

The identification of existing research gaps is crucial for knowledge transfer between fields, particularly for identifying bottlenecks in the receiving field. This issue is discussed in various publications that address different aspects of the problem.

Growing interest in research on energy service companies (ESCOs) has led to an increase in publications over the past decade. Despite this growth, there is still no comprehensive mapping of global research in the field of ESCOs. The purpose of the article [10.1016/j.esr.2024.101516] is to analyze trends in ESCO research, assess the current state of the field, and identify existing gaps in research. This objective is achieved through a systematic literature review and qualitative analysis of recently published articles using bibliometric analysis, co-citation analysis, and keyword analysis to identify recent trends in global ESCO research.

The literature on life cycle assessment (LCA) of lithium-ion batteries (LIB) for transportation applications includes life cycle inventory data relevant to stationary energy storage systems (ESS). However, it does not address the unique characteristics of stationary systems, such as system material balance, operating profiles, and specific end-of-life (EOL) requirements. The literature review [10.1016/j.susmat.2019.e00120] examines existing studies on grid-scale stationary LIB ESS and identifies significant gaps in research related to comprehensive environmental impacts.

The following review [10.1016/j.egyai.2025.100514] examines in detail how reinforcement learning algorithms offer advantages such as fast convergence and stability, particularly relevant to energy management optimization in hybrid electric vehicles. It highlights that deep reinforcement learning outperforms other methods in managing complex energy tasks due to its ability to navigate high-dimensional state spaces. However, challenges remain, such as computational complexity and generalization to different driving conditions.

Comment. To fill the gap in research on electric vehicles, future studies should examine findings from other fields where similar issues have been explored in greater depth, for example, in the development of large language models, and apply this knowledge to the specific field of electric vehicle research.

Setting the research objective

The above publications allow us to formulate a general task that goes beyond the scope of our research. Developing a system for identifying gaps in research based on bibliometric data on a broader topic, such as publications on algorithms from the arxiv.org database, and transferring gaps in knowledge about algorithms to a more specialized field represented in the OnePetro database. The fundamental task of comparing text fields of bibliometric records from different databases is to compile and update a controlled vocabulary. This will be necessary when using VOSviewer for term clustering or GSDMM for text clustering, as well as when searching for co-occurrences of terms in a given window using FP-growth. It is advisable to perform the comparison on a controlled vocabulary, and its composition will significantly affect the results. For example, terms that are irrelevant to the task at hand may behave like noise. Frequently occurring terms may overshadow rare, meaningful, complex terms.

Based on the above, the objective of this study was to take a large dictionary of IEEE controlled terms and analyze how it would affect the formation of thematic clusters when using VOSviewer, the most commonly deployed software for bibliometric analysis. The study included a comparison of results obtained based on different sets of IEEE terms and author keywords in order to determine an approach to using a controlled vocabulary to identify research gaps in a specific context beyond the

scope of IEEE, such as OnePetro, which is planned for the future. But to do this, it is necessary to understand how best to use the terms in the controlled vocabulary.

Bibliometric records exported from the IEEE Xplore database containing the terms “energy,” “systems,” and “technologies” were analyzed. However, the study of the topic of energy system technologies was not part of the main objective of this work and is considered an additional result.

Materials and Methods

The bibliometric data used in this study was exported from the IEEE Xplore open abstract database.

The selection of bibliometric data was carried out in several stages. At the first stage, a list of journals with the largest number of publications related to energy systems technologies was compiled. Twenty-five journals were selected. In the next stage, 10 journals were selected from the 25 that had a high SJR, a high citation index, a sufficient number of publications over three years, and belonged to the category “Energy Engineering and Power Technology”. Then, only bibliometric data from these journals corresponding to the selected topic were exported.

General Characteristics of Journals

Search by query “(“All Metadata”: Energy) OR (“All Metadata”: power) AND (“All Metadata”: system) AND (“All Metadata”: technology); Filters Applied: Journals” provided a list of the top 25 journals included in the IEEE Xplore system with the maximum number of publications on this query (valid as of 15-10-2025).

1. IEEE Access (20,719) not in Energy Engineering and Power Technology Category
2. IEEE Transactions on Vehicular Technology (7,239)
3. Journal of Lightwave Technology (6,065)
4. IEEE Transactions on Power Electronics (5,626)
5. IEEE Transactions on Industrial Electronics (4,367)
6. IEEE Photonics Technology Letters (4,029)
7. IEEE Transactions on Power Systems (3,702)
8. IEEE Transactions on Industry Applications (3,601)
9. IEEE Internet of Things Journal (3,520)
10. IEEE Sensors Journal (3,492)
11. IEEE Transactions on Circuits and Systems II: Express Briefs (2,950)
12. IEEE Transactions on Circuits and Systems I: Regular Papers (2,927)
13. IEEE Transactions on Applied Superconductivity (2,777)
14. IEEE Transactions on Wireless Communications (2,733)
15. IEEE Transactions on Power Delivery (2,715)
16. IEEE Transactions on Instrumentation and Measurement (2,710)
17. IEEE Transactions on Smart Grid (2,438)
18. IEEE Journal of Solid-State Circuits (2,397)
19. IEEE Transactions on Very Large Scale Integration (VLSI) Systems (2,371)
20. IEEE Transactions on Communications (2,212)
21. IEEE Transactions on Microwave Theory and Techniques (2,201)
22. IEEE Transactions on Electron Devices (2,042)
23. IEEE Transactions on Plasma Science (1,951)
24. IEEE Transactions on Industrial Informatics (1,937)
25. IEEE Transactions on Magnetics (1,853)

From this list, 10 journals with the highest SJR values were selected. The journals were chosen based on a search of the titles of 25 journals in the “scimagojr 2024.csv” file, which contains a list of journals, their metadata, and ratings.

The results are presented in Table 1.

Table 1. Top 10 journals with high SJR in the category “Energy Engineering and Power Technology,” selected from a list of 25 journals with the highest number of publications on the topic of energy systems technology.

Title	SJR	Total Docs. (3years)	Citations / Doc. (2years)
IEEE Transactions on Smart Grid	4.608	1272	11.84
IEEE Transactions on Wireless Communications	4.454	2007	12.26
IEEE Transactions on Power Systems	3.629	1506	8.58
IEEE Transactions on Communications	3.492	1676	9.83
IEEE Transactions on Industrial Informatics	3.416	2671	13.30
IEEE Journal of Solid-State Circuits	3.362	935	6.96
IEEE Transactions on Power Electronics	3.083	3889	8.16
IEEE Transactions on Industrial Electronics	3.006	3718	9.67
IEEE Internet of Things Journal	2.483	4971	10.62
IEEE Transactions on Vehicular Technology	2.156	3814	8.24

All journals listed in the table are included in Q1, have a sufficient number of publications over three years, and have a high average citation rate for articles. **Note:** for analyzing a large number of bibliometric records, the average citation rate is a more relevant indicator than the Hirsch index, which characterizes only individual highly cited publications.

Brief Description of Exported Bibliometric Records

For each of the journals listed in Table 1, bibliometric data corresponding to the following types of queries were collected: (“Publication Title”: IEEE Journal Name) AND (“All Metadata”: energy) OR (“All Metadata”: power)) AND (“All Metadata”: system) AND (“All Metadata”: technology) AND (“Publication Year”: Year). Data is current as of October 16, 2025. The results are presented by year: 2023 → 2024 → 2025 for each journal in the following list:

- IEEE Transactions on Smart Grid: 189 → 244 → 253.
- IEEE Transactions on Wireless Communications: 220 → 422 → 334.
- IEEE Transactions on Power Systems: 210 → 275 → 248.
- IEEE Transactions on Communications: 493 → 708 → 903.
- IEEE Transactions on Industrial Informatics: 184 → 293 → 257.
- IEEE Journal of Solid-State Circuits: 115 → 138 → 194.
- IEEE Transactions on Power Electronics: 525 → 561 → 814.
- IEEE Transactions on Industrial Electronics: 187 → 249 → 188.
- IEEE Internet of Things Journal: 461 → 747 → 965.
- IEEE Transactions on Vehicular Technology: 564 → 646 → 845.

A total of 10,993 records were obtained when exporting the bibliometric data. After removing duplicate records, 10,329 remained. Among the records, editorial articles were found and retained because an analysis showed that they contained complete abstract fields.

Record quality: only 3 cells were empty in the *IEEE Terms* field; 44 cells were empty in the *Author Keywords* field; 20 cells were empty in the *Abstract* field. In the *Author Keywords* field, abbreviated titles in parentheses were removed to avoid differences in keywords with and without abbreviations. When constructing the network of term occurrences in the *IEEE Terms* and *Author Keywords* fields, lemmatization was not performed. The reason for this is that it is not done in most published articles. One of the goals of this study was to compare the traditional approach using the

IEEE Terms and *Author Keywords* fields with the proposed options for extracting key terms from title and abstract texts.

About the Controlled Vocabulary

The search for terms in the texts of titles and abstracts was carried out using a controlled vocabulary, which we will also refer to as the *general vocabulary* to distinguish it, for example, from the vocabulary compiled from terms in a specific sample of bibliometric records.

The general dictionary *IEEE Terms* was compiled from the following sources: the IEEE Terms field of data exported in this study, similar data from bibliometric records exported from IEEE Xplore previously used by the author of this study, and IEEE terms extracted from the July 2025 IEEE Thesaurus Version 1.04. After merging, the resulting list of terms was converted to lowercase and lemmatized. Duplicate terms were removed, and manual edits were made to correct errors that occurred when extracting terms from the July 2025 IEEE Thesaurus, Version 1.04. This work used a general dictionary containing 25,639 terms. A dictionary approach was chosen because it is easy to add to and string operations using it are fast. The main difficulty lies in editing dictionaries, particularly in identifying errors. However, this approach is more transparent than using complex libraries, especially those written in Python.

Method for Creating of Terms Co-Occurrence Network

The co-occurrence networks of terms were created using the VOSviewer program [10.1007/s11192-009-0146-3]. The following key term options were used: 1. Author keywords, 2. IEEE terms from this set of bibliometric records, 3. Key terms from the general dictionary found in the texts of titles and abstracts. The latter were used as follows: from the terms found in the texts, a field was constructed similar to the "Index keywords" field in Scopus data. That is, each occurrence of a term was counted once in each record, and repetitions were excluded.

When using dictionary terms, three networks were constructed: 1. terms found throughout the entire IEEE Terms dictionary, 2. only multi-word terms from this dictionary, and 3. mixed occurrence, which took into account all multi-word terms and single-word terms that did not occur in multi-word terms or occurred rarely, in no more than two multi-word terms.

Results and Discussions

Analysis of Author Keywords Co-Occurrence Network

In this section, data from the *Author Keywords* field of all 10,329 bibliometric records were analyzed. The visualization of the co-occurrence map of *Author Keywords* obtained using VOSviewer is shown in Figure 1.

General characteristics of the data used to construct the map: 25,438 — total number of terms; 1,362 — number of terms occurring 5 or more times. Based on this data, 9 clusters were obtained, with cluster 9 containing only 6 terms.

What to pay attention to: Author keywords are more diverse, *but those appearing 5 or more times are even fewer than in IEEE Terms* (see below for a similar analysis for the IEEE Terms field and terms extracted from title texts and abstracts). There were also more clusters than in the other versions of the analysis performed.

It is worth noting that among the author's keywords exceeding the 5-word threshold, there are many multi-word terms that provide a more detailed description of the subject. In contrast, the terms found in titles and abstracts contain many single-word terms, that is, terms that the authors would not have chosen as keywords.

Table 2 presents three fields: keyword → name of the term in the Author Keywords field, occurrences → frequency of occurrence of the term in this field, total link strength → assessment of the connection of this term with others obtained by VOSviewer.

The field header names in the tables closely resemble the headers derived from files that were exported from the program VOSviewer: label → label; occurrences → weight; citation score → score<Avg. norm. citations>.

Table 3. Top 10 terms in the first cluster (red) and their features.

label	occurrences	citation score
electric vehicle	114	1.3995
model predictive control	94	1.2713
permanent magnet synchronous motor	90	1.2677
electric vehicles	81	1.2731
energy management	80	0.868
fault diagnosis	79	1.4944
renewable energy	60	1.3856
stability	54	1.0294
demand response	52	1.8897
microgrid	52	1.0784

The table shows that research in the field of *electric vehicles* dominates both in terms of volume and is well cited. The most cited topic is *demand response* (1.89), and *fault diagnosis* also exceeds its potential (1.49), indicating that narrower, more specialized topics may exceed the main topic of *electric vehicles* in terms of normalized scientific impact.

As already mentioned, the Author Keywords were not lemmatized, which resulted in two terms electric vehicle and electric vehicles.

Table 4. Top 10 terms in the second cluster (green) and their features.

label	occurrences	citation score
reconfigurable intelligent surface	375	1.2286
energy efficiency	277	0.7776
non-orthogonal multiple access	213	0.9993
intelligent reflecting surface	190	1.006
integrated sensing and communication	156	1.8613
power allocation	141	0.953
physical layer security	125	1.0419
outage probability	109	1.1765
beamforming	103	1.4241
nonorthogonal multiple access	95	1.1004

The term *reconfigurable intelligent surface* is the most common (375), but has only an average citation rate (1.23), while *integrated sensing and communication* technologies achieve a much higher normalized impact (1.86). *Energy efficiency*, meanwhile, ranks second in terms of frequency (277), but has a low citation rate of 0.78, while *beamforming* (103, 1.42), *outage probability* (109, 1.18), *physical layer security* (125, 1.04) and *nonorthogonal multiple access* (95, 1.1) show that these issues remain relevant.

Table 5. Top 10 terms in the third cluster (blue) and their features.

label	occurrences	citation score
internet of things	431	0.9057
deep learning	234	1.0358
federated learning	163	1.1088
blockchain	154	1.3165

General characteristics of the data used to construct the map: 2958 — total number of terms; 1379 — number of terms occurring 5 or more times. Based on this data, 8 clusters were obtained, with clusters 5–8 containing only one term.

IEEE Terms from the corresponding field form fewer clusters than the author keywords. The total number of these terms is 8.6 times less than the author's, but the number of terms that occur 5 or more times is even slightly higher: 1379 vs 1362.

The tables below contain the main terms presented in the figure above.

Table 6. Top 10 keywords most frequently found in the EEE Terms field.

keyword	occurrences	total link strength
optimization	1837	12784
internet of things	1440	9794
wireless communication	1345	8946
resource management	1276	8942
voltage control	1126	7720
switches	976	6717
costs	815	5602
training	800	5707
mathematical models	759	5158
array signal processing	745	5123

Optimization dominates both in terms of frequency (1837) and overall connectivity with other terms (12784), acting as a thematic link, while *Internet of Things*, *wireless communication*, *resource management*, and *voltage control* follow it in frequency of occurrence and in connectivity. *Switches*, *mathematical models* and *array signal processing* demonstrate proportionally high connectivity, suggesting that these terms serve as connectors for more narrowly focused technical terms.

This table contains both single-word and multi-word terms. Obviously, these terms were not originally lemmatized, but they have been normalized in terms of spelling. Multi-word terms, such as *array signal processing*, already exist in our dictionary in their lemmatized form *array signal process*. Single-word terms are very general in nature and are also included in the dictionary, for example, *optimization*.

The three clusters with the highest number of terms are presented in the following tables.

Table 7. Top 10 terms in the first cluster (red) and their features.

label	occurrences	citation score
optimization	1837	0.9748
wireless communication	1345	1.0302
resource management	1276	0.9433
array signal processing	745	1.069
sensors	682	1.0379
autonomous aerial vehicles	648	1.1925
interference	541	0.9218
noma	530	0.9903
delays	518	1.0009
vectors	492	0.7956

While high-frequency terms such as *optimization*, *wireless communication*, and *resource management* demonstrate stable but not exceptional citation rates, the most influential terms with the highest score of 1.1925 relate to the less common field of *autonomous aerial vehicles*. Similarly, *array signal processing* and *sensors* show a strong combination of moderate frequency and high impact. In contrast, the category *vectors* stands out for its relatively low frequency and citation rate.

Table 8. Top 10 terms in the second cluster (green) and their features.

label	occurrences	citation score
internet of things	1440	0.9583
costs	815	1.183
training	800	0.9003
computational modeling	709	1.0015
real-time systems	602	0.8249
servers	581	1.0668
energy consumption	573	0.9546
security	560	1.119
accuracy	549	0.7923
task analysis	496	1.3902

Although the term *Internet of Things* ranks first by frequency of occurrence (1,440), its citation rate is average (0.96), while the term *task analysis* (496) has the highest normalized citation (1.39), and *cost* and *security* also exceed it in terms of frequency and citation, indicating that economic and trust issues attract more attention from scientists than general publications on IoT.

Table 9. Top 10 terms in the third cluster (blue) and their features.

label	occurrences	citation score
voltage control	1126	1.0337
switches	976	0.875
topology	723	0.9658
voltage	690	0.8576
capacitors	591	0.8486
batteries	502	1.0635
inverters	489	0.9975
inductors	418	0.8074
couplings	371	1.1872
impedance	366	0.7788

The term *voltage control* dominates this table with 1,126 mentions, with good average citation ratio 1.03, while the much rare term *couplings* (371) and *batteries* (502) has the highest normalized citation rate of 1.19 and 1.06.

In this cluster, the prevalence of single-word terms among frequently occurring terms is particularly noticeable.

Analysis of the Co-Occurrence Network of All Terms from the General Vocabulary

In this section, data from *all terms from the general vocabulary* of all 10,329 bibliometric records were analyzed. The visualization of the co-occurrence map of these terms obtained using VOSviewer is shown in Figure 3.

General characteristics of the data used to construct the map: 6862 — total number of terms; 3326 — number of terms occurring 5 or more times. Based on this data, 4 clusters were obtained, cluster 4 is small.

That is, there are few clusters, and the entries are perceived as fairly homogeneous text with a small number of subtopics.

The three clusters with the highest number of terms are presented in the following tables.

Table 11. Top 10 terms in the first cluster (red) and their features.

label	occurrences	citation score
base	6798	1.019
system	6636	1.0003
method	4502	1.0507
design	3966	0.9907
effective	3913	1.0698
control	2529	1.1435
current	1939	0.9208
simulation	1840	0.942
dynamic	1714	1.0694
voltage	1678	0.9305

The data shows that the most frequently occurring terms, such as *base*, *system*, *method*, and *design* have citation indices close to 1.0. The term *control* has the highest index, at 1.1435.

Table 12. Top 10 terms in the second cluster (green) and their features.

label	occurrences	citation score
model	3789	1.053
time	3388	0.9639
while	2707	0.9504
network	2700	1.0555
enhance	2481	1.0567
data	2457	0.9622
challenge	2431	1.0256
energy	2181	0.9856
framework	1833	1.0185
technology	1705	1.0772

The terms in this cluster are single words, have a similar citation index, and cover topics of a general nature.

Table 13. Top 10 terms in the third cluster (blue) and their features.

label	occurrences	citation score
power	4206	0.9813
problem	3391	1.0445
algorithm	3172	1.0557
compare	2732	0.9819
optimization	2670	1.0265
communication	2242	1.1265
well	1994	1.0667
simulation result	1939	1.0097
rate	1863	0.9505
solve	1819	1.1408

The terms in this cluster are also single words and continue the characteristics of the previous clusters: similar citation frequency and common subject matter.

Table 14. Top 10 keywords most frequently occurring in a field, composed of multi-word terms found in title and abstract texts.

keyword	occurrences	total link strength
simulation result	1939	12047
internet of thing	988	5628
energy consumption	947	5858
energy efficiency	866	5680
base station	717	4863
power system	705	3188
reconfigurable intelligent surface	615	4388
resource allocation	598	4237
unman aerial vehicle	551	3413
wireless communication	510	3284

This table presents a set of specific, relevant keywords. The most prominent term, “simulation result,” is the most common and well linked to others. Internet of Things, base station, power system reflects the infrastructure aspects of the topic, energy consumption and energy efficiency represent economic aspects, while reconfigurable intelligent surface, unmanned aerial vehicle, and wireless communication may be promising technologies for research.

The 10 most frequent multi-word terms from the general vocabulary appear 11,508 times in titles and abstracts (compared to 4,986 times for the 10 most frequent author keywords). Thus, even extracting only the most frequent multi-word terms from the general vocabulary in the texts of titles and abstracts is effective.

The three clusters with the highest number of terms are presented in the following tables.

Table 15. Top 10 terms in the first cluster (red) and their features.

label	occurrences	citation score
power system	705	1.1171
electric vehicle	376	1.2786
wireless power transfer	364	1.0352
renewable energy	281	1.1577
power grid	258	1.2438
power flow	246	0.9994
output power	233	0.8665
power electronic	232	0.9891
energy management	203	1.2006
model predictive control	187	1.3886

The data in the table describes the field of energy research focused on integrating renewable energy sources, electric vehicles, and advanced control systems into the energy system. It is noteworthy that the topics with the highest citation index are *model predictive control* and *electric vehicle*, indicating that algorithms for control and electrification of transport are current areas of research.

Table 16. Top 10 terms in the second cluster (green) and their features.

label	occurrences	citation score
internet of thing	988	0.8913
energy consumption	947	0.9835
resource allocation	598	1.0028
unman aerial vehicle	551	1.2674

deep reinforcement learn	493	1.2409
reinforcement learn	336	1.1355
neural network	333	1.1284
edge compute	316	1.1213
deep learn	293	1.0315
data transmission	266	0.8688

The data in the table shows that frequently used terms such as *Internet of Things*, *energy consumption*, and *data transmission* have a low citation index, indicating that the topics they describe are well known. In contrast, the terms *unmanned aerial vehicles*, *deep reinforcement learning*, and *reinforcement learning* predominate in more frequently cited works. These topics, along with *edge computing*, *neural networks*, and *deep learning*, are becoming hot areas of research.

Table 17. Top 10 terms in the third cluster (blue) and their features.

label	occurrences	citation score
simulation result	1939	1.0097
energy efficiency	866	0.8519
base station	717	1.032
reconfigurable intelligent surface	615	1.2298
wireless communication	510	1.1225
communication system	477	0.8943
power consumption	467	0.697
system performance	391	1.05
channel state information	390	0.9551
phase shift	372	0.9798

The data shows that the research topics of *wireless communication* and *base stations* are reflected in a large number of publications and have a high citation rate. The most interesting topic is the *reconfigurable intelligent surface*. The terms *energy efficiency* and *power consumption* appear more frequently in publications with low citation rates. Publications containing technical terms such as *channel state information* and *phase shift* are less common and have a citation rate closer to average.

Analysis of the Co-Occurrence Network of Mixed Terms from the General Vocabulary

At this stage, an attempt has been made to reduce the number of single-word terms in the general dictionary. Only single-word terms that are not part of multi-word terms or are part of multi-word terms but no more than twice have been retained. For this option, the concept of *mixed terms* is used.

In this section, data from *mixed terms from the general vocabulary* of all 10,329 bibliometric records were analyzed. The visualization of the co-occurrence map of these terms obtained using VOSviewer is shown in Figure 5.

At this stage, an attempt has been made to reduce the number of single-word terms in the general dictionary. Only terms that are not part of multi-word terms or are part of them but no more than twice have been retained.

General characteristics of the data used to construct the map: 6827 — total number of terms; 3306 — number of terms occurring 5 or more times. Based on this data, 4 clusters were obtained, cluster 4 is small.

This option is minimally distinct from the alternative that employed all terms from the general dictionary. So, this section contains only tables, excluding comments.

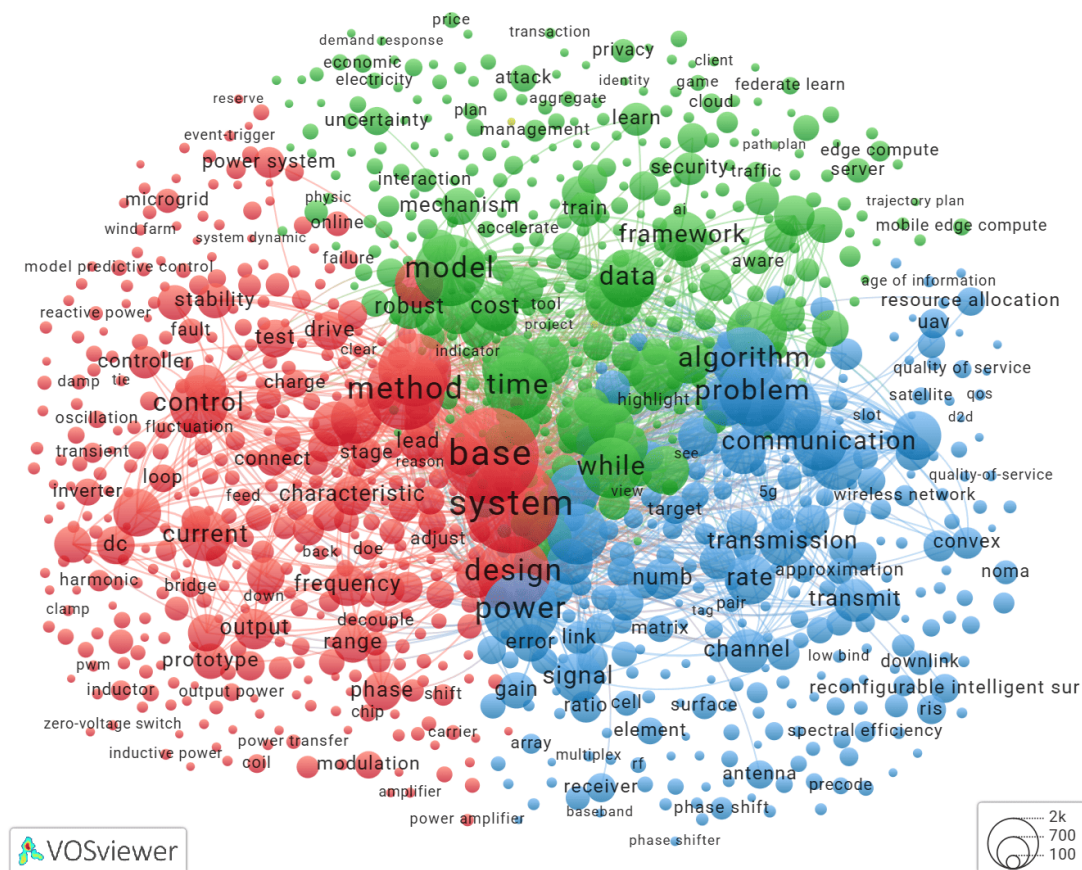


Figure 5. A co-occurrence map of mixed terms from the general vocabulary, illustrating the structure of research in the field of energy systems technologies for 2023-2025.

The tables below contain the main terms presented in the figure above.

Table 18. Top 10 keywords most frequently occurring in a field, composed of mixed terms found in title and abstract texts.

keyword	occurrences	total link strength
base	6798	217921
system	6636	213789
method	4502	142463
power	4206	139009
design	3962	129139
effective	3913	125658
model	3789	120839
problem	3391	115384
time	3388	111217
algorithm	3170	108159

The terms in this table and in the table compiled for all IEEE terms are identical. The difference in overall connection strength between the two tables can be explained by the different sets of terms. However, the slight difference in frequency of the terms “algorithm” and “design” (3,170 vs. 3,172 and 3,962 vs. 3,966) is difficult to explain. It cannot be ruled out that this difference is due to the optimization of the search algorithm for terms of different lengths, which was implemented in a specific version of a grep-type utility. Typically, sorting is done by number of characters rather than

number of words. Another possible reason is that incorrect strings were omitted from the dictionary. The current dictionary consists of 25,639 strings, so some inaccuracies are possible; for example, an unseparated space or short dash may be used instead of a hyphen. It is also worth noting that this data was taken from files generated by VOSviewer. The noted difference did not affect the conclusions made in this study but requires a separate technical analysis to identify possible causes. The dictionary is regularly updated with new terms and checked for inaccuracies.

Note: Testing the use of the dictionary with various text utilities is an important task that must be performed regularly. The most obvious approach to testing is to use comparison programs such as WinMerge, which make it easy to track changes made by utilities.

As mentioned earlier, we will supply the tables with no commentary.

Table 19. Top 10 terms in the first cluster (red) and their features.

label	occurrences	citation score
base	6798	1.019
system	6636	1.0003
method	4502	1.0507
design	3962	0.991
effective	3913	1.0698
control	2529	1.1435
current	1939	0.9208
simulation	1840	0.942
dynamic	1714	1.0694
voltage	1678	0.9305

Table 20. Top 10 terms in the second cluster (green) and their features.

label	occurrences	citation score
model	3789	1.053
time	3388	0.9639
while	2707	0.9504
network	2700	1.0555
enhance	2481	1.0567
data	2457	0.9622
challenge	2431	1.0256
energy	2181	0.9856
framework	1833	1.0185
technology	1705	1.0772
process	1633	1.0394

Table 21. Top 10 terms in the third cluster (blue) and their features.

label	occurrences	citation score
power	4206	0.9813
problem	3391	1.0445
algorithm	3170	1.0556
compare	2732	0.9819
optimization	2670	1.0265
communication	2242	1.1265
well	1994	1.0667
simulation result	1939	1.0097
rate	1863	0.9505
solve	1819	1.1408

In all three of the tables above, only one term turned out to be two words: *simulation result*.

Thus, the analysis of the occurrence of mixed terms in the texts of titles and abstracts shows that single-word terms will prevail, even if they are absent or rarely occur as components of multi-word terms. Therefore, when using a dictionary of common technical terms, it is justified to include only those single-word terms that can denote a specific subject area, for example, *blockchain*, excluding those that rarely appear in other compound terms; such an exclusion is justified since it does not guarantee the frequent occurrence of such terms in written works.

Conclusions

The simplest conclusion that can be drawn from the analysis above is that the most comprehensive dictionary of terms will not necessarily be the most effective for identifying topics of interest. It can be preliminarily stated that it is not advisable to limit oneself to technical means of dictionary selection. Compiling a dictionary significantly simplifies the work of an expert analyzing their area of interest, but the dictionary should be peer-reviewed to better suit the task being solved. Problem formulation is more of a professional experience, while analytics only provides arguments to support the validity of the choice made. The hope for a self-sufficient analytical method is more of a myth than a reality. Analytics only assists in making an informed decision; it does not replace the expertise.

References

1. Pavan L. Detecting Main Topics using Dictionary-based Topic Analysis. *IJLLT* 2022;5:48–52. <https://doi.org/10.32996/ijllt.2022.5.12.6>.
2. Carley K. Extracting culture through textual analysis. *Poetics* 1994;22:291–312. [https://doi.org/10.1016/0304-422X\(94\)90011-6](https://doi.org/10.1016/0304-422X(94)90011-6).
3. Cunningham E, Greene D. Knowledge transfer, knowledge gaps, and knowledge silos in citation networks. *PLoS One* 2025;20:e0329302. <https://doi.org/10.1371/journal.pone.0329302>.
4. Small H. Maps of science as interdisciplinary discourse: co-citation contexts and the role of analogy. *Scientometrics* 2010;83:835–49. <https://doi.org/10.1007/s11192-009-0121-z>.
5. Aman V, Gläser J. Investigating Knowledge Flows in Scientific Communities: The Potential of Bibliometric Methods. *Minerva* 2025;63:155–82. <https://doi.org/10.1007/s11024-024-09542-2>.
6. Moradi B, Hermansson L, Ellingsen T, Ask KK, Boisjolly ED, Stein-Beldring EJ, et al. Reshaping the Oil & Gas Industry: The Rise of the Digital Petroleum Engineer. *SPE Offshore Europe Conference & Exhibition, Aberdeen, Scotland, UK: SPE; 2025, p. D021S009R006*. <https://doi.org/10.2118/226792-MS>.
7. Peter Kenneth Minnoh. INTEGRATING MACHINE LEARNING AND PROCESS SYSTEMS ENGINEERING FOR SUSTAINABLE OPTIMIZATION OF PETROLEUM AND PETROCHEMICAL OPERATIONS IN THE U.S. ENERGY SECTOR. *Epra* 2025:1–9. <https://doi.org/10.36713/epra24128>.
8. Tariq Z, Mahmoud M, Abdurraheem A. Machine Learning-Based Improved Pressure–Volume–Temperature Correlations for Black Oil Reservoirs. *Journal of Energy Resources Technology* 2021;143:113003. <https://doi.org/10.1115/1.4050579>.
9. Chen Y, Zhao Y, Kjenner S, Hasan K. Managing Knowledge Transfer in Innovative Complex Systems Development: Case Study of Renewable Energy Project in the Oil and Gas Industry. *INCOSE International Symp* 2023;33:1173–87. <https://doi.org/10.1002/iis2.13076>.
10. Cao C, Xue X, Zhang K, Song L, Zhang L, Yan X, et al. Competitive Knowledge Transfer–Enhanced Surrogate-Assisted Search for Production Optimization. *SPE Journal* 2024;29:3277–92. <https://doi.org/10.2118/219732-PA>.
11. Ferrer-Serrano M, Fuentelsaz L, Latorre-Martínez MP. Knowledge Transfer and Networks: A Bibliometric Approach Through Performance Analysis, Science mapping, and Dynamic Network Analysis. *J Knowl Econ* 2025. <https://doi.org/10.1007/s13132-025-02814-6>.
12. Bengoa A, Maseda A, Iturralde T, Aparicio G. A bibliometric review of the technology transfer literature. *J Technol Transf* 2021;46:1514–50. <https://doi.org/10.1007/s10961-019-09774-5>.

13. Figueiredo N, Patrício L, Pinheiro P. Unveiling university-industry knowledge transfer: insights from bibliographic coupling analysis. *VINE Journal of Information and Knowledge Management Systems* 2025;55:1604–28. <https://doi.org/10.1108/VJKMS-07-2024-0270>.
14. Rani N, Yaqub MZ, Singh N, Magliocca P. Exploring the significance of knowledge transfer for facilitating cross-border acquisitions: an extensive examination of current themes, gaps, and potential future research directions. *JKM* 2025;29:837–69. <https://doi.org/10.1108/JKM-04-2024-0494>.
15. Gu Z, Meng F, Farrukh M. Mapping the Research on Knowledge Transfer: A Scientometrics Approach. *IEEE Access* 2021;9:34647–59. <https://doi.org/10.1109/ACCESS.2021.3061576>.
16. Cebekhulu BMB, Mathaba TND, Mbohwa C. Identifying trends and research gaps in ESCO research: A systematic literature review. *Energy Strategy Reviews* 2024;55:101516. <https://doi.org/10.1016/j.esr.2024.101516>.
17. Pellow MA, Ambrose H, Mulvaney D, Betita R, Shaw S. Research gaps in environmental life cycle assessments of lithium ion batteries for grid-scale stationary energy storage systems: End-of-life options and other issues. *Sustainable Materials and Technologies* 2020;23:e00120. <https://doi.org/10.1016/j.susmat.2019.e00120>.
18. Boukoberine MN, Zia MF, Berghout T, Benbouzid M. Reinforcement learning-based energy management for hybrid electric vehicles: A comprehensive up-to-date review on methods, challenges, and research gaps. *Energy and AI* 2025;21:100514. <https://doi.org/10.1016/j.egyai.2025.100514>.
19. Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 2010;84:523–38. <https://doi.org/10.1007/s11192-009-0146-3>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.